

5'-UTR 序列特征对 mRNA 丰度的调控作用及优化策略

梅子轮^{1,2}, 张金鹏^{1,2}, 邵佳怡^{1,2}, 徐国强^{1,2}, 任家卫^{1,2}, 张晓梅³, 李会³, 史劲松³, 张晓娟^{1,2*}, 许正宏⁴

1 江南大学 生物工程学院, 江苏 无锡

2 江南大学, 粮食发酵与食品生物制造国家工程研究中心, 江苏 无锡

3 江南大学 生命科学与健康工程学院, 江苏 无锡

4 四川大学 轻工科学与工程学院, 四川 成都

梅子轮, 张金鹏, 邵佳怡, 徐国强, 任家卫, 张晓梅, 李会, 史劲松, 张晓娟, 许正宏. 5'-UTR 序列特征对 mRNA 丰度的调控作用及优化策略[J]. 微生物学报, 2025, 65(8): 3567-3582.

MEI Zilun, ZHANG Jinpeng, SHAO Jiayi, XU Guoqiang, REN Jiawei, ZHANG Xiaomei, LI Hui, SHI Jinsong, ZHANG Xiaojuan, XU Zhenghong. Regulatory effects of 5'-UTR sequence characteristics on mRNA abundance and optimization strategies[J]. *Acta Microbiologica Sinica*, 2025, 65(8): 3567-3582.

摘要:【目的】转录是基因表达的第一步, mRNA 丰度在一定程度上决定着最终蛋白质的表达丰度。近期一些研究发现, 5'-UTR 中不同的核糖体结合位点(ribosome-binding site, RBS)会对下游基因的 mRNA 丰度产生影响。从 mRNA 降解过程的调控因素角度分析, 这种影响可能源自 Shine-Dalgarno (SD)序列与核糖体的结合强度以及 5'-UTR 局部二级结构。【方法】构建了一个包含 518 个突变体的 5'-UTR 突变文库, 结合高通量测序技术, 高效采集各 5'-UTR 突变体对应其调控的下游 *egfp* 基因的 mRNA 丰度, 并通过 RT-qPCR 验证了其有效性。【结果】将各 mRNA 突变体丰度与其对应的 5'-UTR 序列特征进行关联分析, 结果表明, 与核糖体结合强度为中等偏强的 SD 序列最有利于维持高 mRNA 丰度, 结合强度过高或过低均会导致 mRNA 丰度的降低; 完全保守的核心 SD 序列(GGAGG)是保证较高结合强度的关键, 其保守性下降将导致 mRNA 丰度明显下降。当不同序列的 5'-UTR 中 SD 序列接近, 即与核糖体结合强度相似时, 其局部二级结构越不稳定, 对应的 mRNA 丰度越高。【结论】本研究解析了 5'-UTR 的各序列特征对 mRNA 丰度的调控作用, 初步建立了其相互之间的定性模型, 为代谢工程及基因线路中调控元件的理性设计提供了重要参考。

关键词: 5'-UTR; SD 序列; 局部二级结构; mRNA 丰度

资助项目: 国家重点研发计划(2024YFA0918000); 国家自然科学基金(32171421)

This work was supported by the National Key Research and Development Program of China (2024YFA0918000) and the National Natural Science Foundation of China (32171421).

*Corresponding author. E-mail: zhangxj@jiangnan.edu.cn

Received: 2025-01-03; Accepted: 2025-03-23; Published online: 2025-06-06

Regulatory effects of 5'-UTR sequence characteristics on mRNA abundance and optimization strategies

MEI Zilun^{1,2}, ZHANG Jinpeng^{1,2}, SHAO Jiayi^{1,2}, XU Guoqiang^{1,2}, REN Jiawei^{1,2}, ZHANG Xiaomei³, LI Hui³, SHI Jinsong³, ZHANG Xiaojuan^{1,2*}, XU Zhenghong⁴

1 School of Biotechnology, Jiangnan University, Wuxi, Jiangsu, China

2 National Engineering Research Center of Cereal Fermentation and Food Biomanufacturing, Jiangnan University, Wuxi, Jiangsu, China

3 School of Life Sciences and Health Engineering, Jiangnan University, Wuxi, Jiangsu, China

4 College of Biomass Science and Engineering, Sichuan University, Chengdu, Sichuan, China

Abstract: [Objective] Transcription is the first step in gene expression, and the mRNA abundance to a certain extent determines the final protein expression abundance. Recent studies have found that different ribosome-binding sites (RBSs) located in the 5' untranslated region (5'-UTR) can affect the mRNA abundance of the downstream gene. From the perspective of regulatory factors in the mRNA degradation process, the effect may be attributed to the binding strength between the Shine-Dalgarno (SD) sequence and the ribosome and the local secondary structure of the 5'-UTR. **[Methods]** We constructed a 5'-UTR mutant library with a size of 528. High-throughput sequencing was employed to efficiently collect the information on the mRNA abundance of downstream *egfp* corresponding to various 5'-UTR variants. The effectiveness was verified by RT-qPCR. **[Results]** The association between abundance of each mRNA mutant and its corresponding 5'-UTR sequence was analyzed. The results showed that the SD sequence with moderate to strong binding strength to the ribosome was most conducive to maintaining high mRNA abundance. Too high or low binding strength will lead to a reduction in the mRNA abundance. The completely conserved core SD sequence (GGAGG) was the key to ensuring high binding strength, and the decline in conservation would cause a significant decrease in the mRNA abundance. When the SD sequence was similar among different 5'-UTR variants, i.e., the binding strength of the SD sequence to the ribosome was comparable, the local secondary structure of the 5'-UTR was instable and the abundance of corresponding mRNA was high. **[Conclusion]** This study delves into the regulatory effects of 5'-UTR sequence features 5'-UTR on the mRNA abundance and establishes a qualitative model of their interrelationships, providing a reference for the rational design of regulatory elements in metabolic engineering and gene circuits.

Keywords: 5'-UTR; SD sequence; local secondary structure; mRNA abundance

基因表达主要分为转录和翻译 2 个步骤，具体包括(1) mRNA 合成；(2) 蛋白质合成；(3) mRNA 降解；(4) 蛋白质降解，其中的每一步都可用于调控基因表达^[1]。在代谢工程中，基因表

达调控元件已被广泛用于调节代谢通量^[2]，常见的调控元件包括启动子、核糖体结合位点(ribosome-binding site, RBS)、终止子和转录因子等。其中，RBS 可用于调控基因表达的翻译过

程^[3-4], 例如, 调控操纵子内部基因的翻译效率, 平衡多酶系统, 避免中间产物的积累^[5]。RBS 位于 mRNA 的 5'-非翻译区(5'-untranslated region, 5'-UTR), 由 Shine-Dalgarno (SD) 序列及其与起始密码子之间的间隔区域组成^[6-7]。核糖体小亚基的 16S rRNA 的 3'末端(主要为一段 9 nt 序列: 5'-ACCUGCUUA-3', 称为 anti-SD 序列)可以识别并结合 SD 序列, 从此在起始密码子附近招募核糖体复合物并开启翻译过程^[6,8]。

在过去对 RBS 的研究和应用中研究者大多关注通过改变 RBS 来调控翻译过程, 从而影响蛋白质的表达量^[5,9]。然而, 最近的研究发现, 在不同 RBS 调控下, 下游基因的 mRNA 丰度也发生了显著变化。Duan 等^[10]通过改变 RBS 与 anti-SD 互补配对的碱基数量以及间隔区域的碱基组成, 使下游 PMK-gfp 蛋白基因的 mRNA 丰度上调了 5.9 倍; Cetnar 等^[11]在使用通过计算得到不同翻译起始速率的 RBS 调控下游 *sfGFP* 表达后, 其 mRNA 丰度上调了 7.85 倍。由此可见, 不同的 RBS 序列不仅调控翻译水平, 还会显著影响基因的 mRNA 丰度。研究表明, 尽管细菌基因表达涉及多步反应, 但 mRNA 丰度对最终蛋白质浓度有重要影响^[12-13]。mRNA 丰度由 mRNA 合成和降解 2 个过程共同调控, 其中 mRNA 合成主要由启动子和转录因子调控, 而 mRNA 降解则主要受到 RNA 降解酶的作用影响。大肠杆菌体内存在多种 RNA 降解酶, 包括内切酶(RNase E、G、III)、外切酶(RNase II、PNPase、RNase R)以及一些协同作用的辅助酶, 它们通过多种方式识别 mRNA 并与之结合, 进而发挥降解功能。这些酶在多途径过程中共同作用, 决定了 mRNA 的降解速度^[11]。

影响原核生物 mRNA 稳定性的因素众多, 包括翻译速率、二级结构、转录速度等^[14]。Pedersen 等^[15]通过在前 200 个密码子的编码区域插入不同特异性翻译速率的密码子, 调控核糖体在 mRNA 上的分布间距, 发现更稳定的 mRNA 在起始部分具有更高的核糖体密度。

Richards 等^[16]探究了核糖体结合和位移对 RNase E 2 种降解机制的影响: 通过 mRNA 单磷酸化的 5'端内部(5'端依赖型降解)或直接进入 mRNA 内部(非 5'端依赖型降解), 证明了这 2 种降解方式均可被在翻译起始位点结合的核糖体阻碍, 且对 5'端依赖型降解的抵抗作用更显著; 而翻译延伸过程中核糖体可以有效阻止 RNase E 从内部降解 mRNA。Carrier 等^[17]在 mRNA 的 5'-UTR 内设计不同特定自由能的二级结构, 结果对 mRNA 的半衰期产生了 10 倍以上的差异影响。简单来说, 当 mRNA 对这些降解酶的可及性降低时, 无论是通过覆盖在其上的核糖体复合物^[18], 还是通过 mRNA 本身的高级结构^[19], 其降解速度将会减缓, 从而带来丰度的提高。RBS 的强弱^[13](针对翻译起始速率)、5'-UTR 区域的二级结构稳定性^[20]、5'-UTR 的碱基组成^[19]等可以直接或间接地影响 mRNA 稳定性, 因此可从这些序列特征中寻找 5'-UTR 调控 mRNA 丰度的主要因素。

准确获取基因型和表型的相关数据并构建两者之间的对应关系, 是解析调控元件构效关系的前提。随着测序技术的发展, 越来越多高效精准的方法被建立。高通量测序技术[也称下一代测序技术(next-generation sequencing, NGS)]通过并行测序的方式可以同时测序数百万个 DNA 片段。利用高通量测序的高效和高分辨率特性可在单次实验中实现对目标区域不同序列变体调控效果的系统定量解析。2009 年, Patwardhan 等^[21]首次提出利用大规模并行 DNA 合成和测序对调控元件序列进行单核苷酸分辨率的高通量分析的方法; 他们构建了启动子文库并在体外进行转录, 通过对模板 DNA 以及 mRNA (逆转录为 cDNA)进行高通量测序, 计算得到每个模板 DNA 的转录本数量, 以此来表征不同启动子的转录活性。Kosuri 等^[13]使用该方法实现了对 114 个启动子和 111 个 RBS 的所有组合调控下大肠杆菌体内目的基因 mRNA 水平的定量表征。因此可利用高通量测序技术, 分

别对含目标 5'-UTR 序列的 DNA 及 mRNA 文库进行测序, 获取各变体调控下的 mRNA 相对丰度, 以实现 5'-UTR 调控 mRNA 丰度的系统解析。

目前, 关于 5'-UTR 对基因表达转录水平的影响缺乏系统研究, 其背后的影响因素以及构效关系尚不明确。基于此, 本研究设计并构建了 5'-UTR 突变文库, 在 SD 序列区域、SD 序列前 2 个碱基处以及间隔区域内分别引入了 3 个、2 个碱基饱和突变以及富含 AC/GT 2 种设计, 共产生了 518 个 5'-UTR 变体序列用于调控下游 *egfp* 基因表达, 结合高通量测序技术检测出下游 *egfp* 基因的 mRNA 丰度。将各 5'-UTR 突变体的序列特征与 mRNA 丰度进行关联分析, 结果表明转录水平与 SD 序列与核糖体的结合强度存在显著关联, 且在相同结合强度下, 5'-UTR 区域的二级结构表现出对转录水平的微弱影响。本研究结果为理性设计元件调控基因表达的丰度提供了一个设计原则。

1 材料与方法

1.1 菌株和质粒

大肠埃希氏菌(*Escherichia coli*) JM109(DE3): *recA1*⁻、*endA1*⁻、*gyrA96*⁻、*thi-1*⁻、*hsdR17* (*rk⁻mk⁺*)、*el4⁻* (*mcrA⁻*) *supE44*⁻、*relA1*⁻、 Δ (*lac-proAB*)/F'[*traD36*、*proAB⁺*、*lacI^f*、*lacZ* Δ M15]用于质粒构建和作为表达宿主。本研究基于 pDXW-13-tac (*E. coli*-*B. flavum* 穿梭探针质粒, 抗性基因为 *kan^r* 和 *cat^r*, 报告基因为 *egfp*, 启动子为诱导性启动子 tac, 来源于本实验室)构建了 pDXW-13-tac-*Bsa* I 质粒(在报告基因前增加 2 个反向互补 *Bsa* I 酶切位点)用于快速替换 RBS。

1.2 主要试剂和仪器

IPTG、卡那霉素、溶菌酶, 生工生物工程(上海)股份有限公司; DNA marker、DNA 聚合酶 Phanta 2×Mix Master Max、DNA 聚合酶 2×*Rapid Taq* Master Mix、限制酶 *Hind* III、限制

酶 *Xho* I、磷酸激酶和 T4 DNA 连接酶、超级感受态制备试剂盒, TaKaRa 公司; 限制酶 *Bsa* I-HF[®] v2, New England Biolabs 公司; 质粒小量提取试剂盒、琼脂糖凝胶 DNA 回收试剂盒、柱式 DNA 纯化试剂盒, 上海捷瑞生物工程有限公司; 总 RNA 提取试剂盒、逆转录试剂盒、染料法定量 PCR 检测试剂盒, 南京诺唯赞生物科技股份有限公司; SuperScript[™] III 反转录酶, ThermoFisher Scientific 公司。

小型离心机, Eppendorf 公司; 紫外可见分光光度计, 上海美谱达仪器有限公司; NanoDrop 微量核酸蛋白浓度测定仪, ThermoFisher Scientific 公司; 核酸电泳仪、PCR 仪、荧光定量 PCR 仪, Bio-Rad 公司; SPX 生化培养箱, 上海跃进医疗器械有限公司; DHZ-DA 型恒温摇床, 太仓市实验设备厂; 超净工作台, 吴江市金晓空调净化有限公司。

1.3 培养基

LB 液体培养基(g/L): 蛋白胨 10.0, 酵母粉 5.0, NaCl 10.0。固体培养基加入琼脂粉 20.0 g/L。

1.4 5'-UTR 文库的构建

5'-UTR 文库(NNNAAGGANM₆/K₆, CACAA GNAGM₆/K₆)是根据大肠杆菌中 16S rRNA 基因 5'-ACCUCUUUA-3'的 3'末端设计的, 其中“N”代表 A/T/C/G。由于 RBS 富含 A 或 G, SD 设计为 AAGGA, 其与 16S rRNA 基因的 3'末端的碱基互补配对确保翻译^[6,22]。此外, 本研究将 RBS 和起始密码子之间的核苷酸设计为 6 个连续的 AC 碱基或 GT 碱基。通过在引物上引入上述序列, 并在引物 5'端添加 *Xho* I 和 *Hind* III 酶切位点及保护碱基, 4 个正向引物和一个反向引物设计见表 1, 引物由亦欣生物科技无锡有限公司合成。使用 DNA 聚合酶以正向引物和反向引物扩增报告基因 *egfp*, 4 类 PCR 产物经 1% 琼脂糖凝胶电泳切胶回收纯化并测定浓度, 胶回收按照琼脂糖凝胶 DNA 回收试剂盒说明书进行。将

表1 构建5'-UTR文库使用的引物

Table 1 Primers used in constructing 5'-UTR library

Primers name	Primer sequences (5'→3')
4N-AC-gfp-F	CCGAAGCTTTAACTAACTAANNNAAGGANACACACATGGTGAGCAAGGGCGAGGAGC
4N-GT-gfp-F	CCGAAGCTTTAACTAACTAANNNAAGGANGTGTGTATGGTGAGCAAGGGCGAGGAGC
1N-AC-gfp-F	CCGAAGCTTTAACTAACTAACACAAGNAGACACACATGGTGAGCAAGGGCGAGGAGC
1N-GT-gfp-F	CCGAAGCTTTAACTAACTAACACAAGNAGGTGTGTATGGTGAGCAAGGGCGAGGAGC
gfp-R	CCGGCTCGAGTTACTTGTACAGCTCGTC

4 类 PCR 回收产物及 pDXW-13-tac 质粒用 *Xho* I 和 *Hind* III 消化后过柱纯化分别得到 4 类目的片段和线性载体并测定浓度, 纯化按照柱式 DNA 纯化试剂盒说明书进行。分别取 4 类目的片段和线性载体各 0.5 pmol 和 0.05 pmol (摩尔比 10:1), 加入 10 U/ μ L T4 DNA 连接酶 2 μ L, 2 $\times$ 连接缓冲液 2 μ L, 超纯水补齐至 20 μ L, 混合后置于 16 $^{\circ}$ C 下反应 12 h, 获得连接产物 1N-AC、4N-AC、1N-GT、4N-GT。将连接产物 1N-AC 和 4N-AC 按体积比 1:64 混合, 连接产物 1N-GT 和 4N-GT 按体积比 1:64 混合, 获得突变文库质粒 AC 库和 GT 库, 然后取突变文库质粒 AC 库、GT 库各 10 μ L 分别化学转化到 *E. coli* JM109(DE3)感受态细胞中, 于 50 mL 含卡那霉素的 LB 液体培养基中 37 $^{\circ}$ C、220 r/min 培养过夜, 所得菌液用于后续诱导表达培养及测序文库构建。

1.5 重组质粒的构建

采用了 Golden Gate Assembly 的方法^[23]构建重组质粒, 该方法在本研究中有 2 个优势: (1) 该方法可以实现片段与载体间的无缝组装; (2) 该方法可以简便高效地在质粒上固定位置替换不同序列。

1.5.1 双反向 *Bsa* I 位点质粒的构建

采用反向 PCR 的方法在报告基因的起始密码子前插入一段包含 2 个反向的 *Bsa* I 酶切位点的序列。通过在插入位点设计 2 条具有重叠部分的特异性引物反向扩增 pDXW-13-tac 质粒骨架, 重叠部分即为插入序列, 引物设计为 rPCR-

Bsa I-F (5'-CGCTagagaccGCTTTCAGATCTGTAACTTGTggtctcgGTGAGCAAGGGCGAGGAGCTG-3')和 rPCR-*Bsa* I-R (5'-cgagaccACAAGTTACAGATCTGGAAAGCggtctctAGCGAATTGTTATCGCTCACAATTCCACAC-3')。PCR 反应体系 (20 μ L): 2 $\times$ Phanta Max Mix (p515) 10 μ L, 上、下游引物 (20 μ mol/L) 各 1 μ L, DNA 模板 1 μ L, ddH₂O 7 μ L。PCR 反应条件: 95 $^{\circ}$ C 预变性 5 min; 95 $^{\circ}$ C 变性 30 s, 57 $^{\circ}$ C 退火 30 s, 72 $^{\circ}$ C 延伸 4 min, 35 个循环; 72 $^{\circ}$ C 终延伸 10 min。然后取 PCR 产物 10 μ L 化学转化入 JM109(DE3)感受态细胞中, 并涂布于含有卡那霉素的 LB 固体培养基上, 37 $^{\circ}$ C 培养过夜, 利用 *E. coli* JM109(DE3)自身的重组能力环化线型质粒。将单菌落挑菌测序, 测序结果符合预期的对应菌落携带质粒命名为 pDXW-13-tac-*Bsa* I。

1.5.2 5'-UTR 的连接

5'-UTR 以引物对的方式(正向和反向引物反向互补)合成。其中, 正向引物的黏性末端设计为 5'-CGCT, 反向引物的黏性末端由融合蛋白的前 4 个核苷酸决定。例如, 当 *egfp* 的前 4 个核苷酸为 GTGA 时, 将反向引物的黏性末端设计为 5'-TCAC, 实现 RBS 与报告基因的无缝组装。退火 PCR 反应体系 (20 μ L): 正、负链引物 (20 μ mol/L) 各 10 μ L。退火 PCR 反应条件: 95 $^{\circ}$ C 预变性 5 min; 降低 0.1 $^{\circ}$ C 8 s, 700 个循环; 72 $^{\circ}$ C 终延伸 10 min。磷酸化 PCR 反应体系 (50 μ L): 磷酸激酶 1 μ L, 退火产物 20 μ L, 10 $\times$ T4 PNK buffer 5 μ L, ATP (100 mmol/L) 0.5 μ L, ddH₂O 23.5 μ L。磷酸化 PCR 反应条件: 37 $^{\circ}$ C

30 min, 65 °C 20 min。磷酸化的 RBS 与 pDXW-13-tac-Bsa I 质粒通过边酶切边连接的方式进行重组连接(图 1): 取目的片段 2 pmol, 载体质粒 0.025 pmol, 加入 20 U/μL T4 DNA 连接酶 1 μL, 10 U/μL Bsa I 限制酶 1 μL, 2×连接缓冲液 1 μL, 超纯水补齐至 10 μL。PCR 反应程序: 37 °C 5 min, 60 °C 5 min。取连接产物 10 μL 化学转化入 *E. coli* JM109(DE3)感受态细胞中, 并涂布于含有卡那霉素的 LB 固体培养基上, 37 °C 培养过夜。挑单菌落进行测序, 测序结果符合预期的对应菌落携带目标重组质粒。

1.6 *egfp* 的 mRNA 丰度测定

将经测序正确的重组菌株接种至装有 2 mL 含有卡那霉素的 LB 液体培养基的 24 孔板中, 37 °C、220 r/min 培养过夜。然后以 2% 的接种体积分数转接到装有 3 mL 含有卡那霉素的 LB 液体培养基的另一个 24 孔板中, 37 °C、220 r/min 培养 1.5 h 至 OD_{600} 为 0.4–0.6。随后以终浓度为 0.4 mmol/L 的 IPTG 于 37 °C、220 r/min 诱导培养 2 h。取 1 mL 菌液按照总 RNA 提取试剂盒说明书提取总 RNA, 然后按照逆转录及 qPCR 预混液试剂盒说明书将 RNA 逆转录为 cDNA。qPCR 操作步骤参考染料法定量 PCR 检测试剂盒说明书, 选用 16S rRNA 基因作为内参基因, 使用 $2^{-\Delta\Delta C_t}$ 的方法^[24]分析 mRNA 丰度, 如公

式(1)、(2)所示。以数值大于 1 表示 mRNA 丰度上调, 小于 1 且大于 0 表示 mRNA 丰度下调。

$$\Delta C_{t,x} = C_{t,x} - C_{t,x,r} \quad (1)$$

式中: $C_{t,x}$ 为样本目的基因的 C_t 值; $C_{t,x,r}$ 为样本内参基因的 C_t 值。

$$\text{mRNA level} = 2^{-(\Delta C_{t,x} - \Delta C_{t,xr})} \quad (2)$$

式中: $\Delta C_{t,x}$ 为样本 x 的 ΔC_t 值; $\Delta C_{t,xr}$ 为对照样本的 ΔC_t 值; 计算所得的 mRNA level 用以表征 mRNA 丰度, 其数值代表样本 x 相对对照样本的 mRNA 丰度的比值。

1.7 高通量测序

将所获得的 2 类携带突变文库质粒的菌群按照 1.6 节的方法进行诱导培养(其中转接培养基体积为 4 mL)。取 1.5 mL 菌液装于 2 mL 无菌 EP 管中置于 -80 °C 冰箱保存, 用于后续试验操作。另取 1.5 mL 菌液按照质粒小量提取试剂盒说明书提取质粒, 并通过 PCR 扩增从目的报告基因所在表达盒转录起始位点开始的 200 bp 序列。经 1% 琼脂糖凝胶电泳切胶回收纯化并测定浓度后, 将 AC 库和 GT 库扩增出的序列按摩尔比 1:1 混合, 得到总 DNA 突变测序文库。取出单独保存的 1.5 mL 菌液, 按照总 RNA 提取试剂盒说明书提取总 RNA, 使用 SuperScript™ III 反转录酶从目的 mRNA 距离转录起始位点 200 bp 处开始逆转录出 cDNA, 然后通过 PCR 扩增

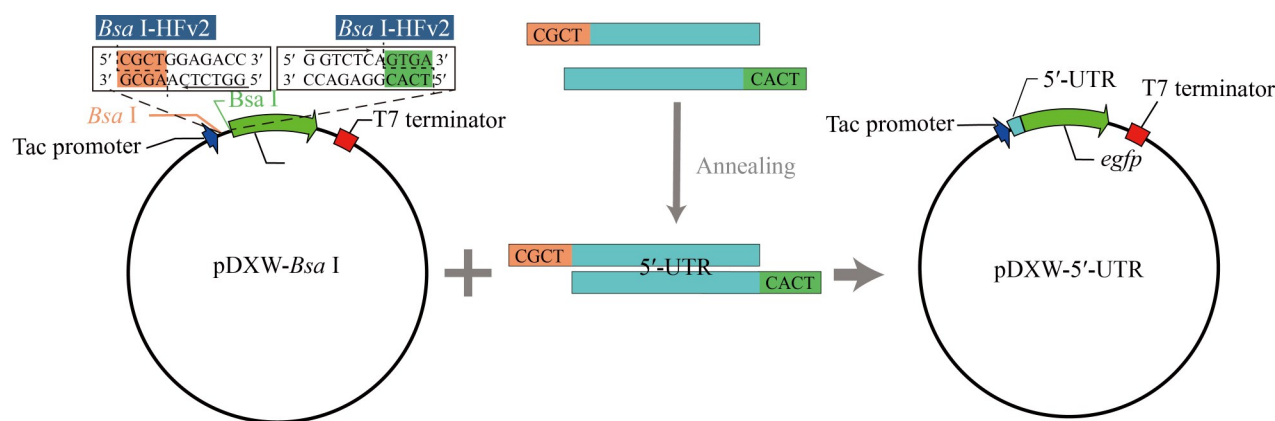


图1 5'-UTR连接示意图

Figure 1 Schematic of 5'-UTR ligation.

cDNA 片段, 经 1% 琼脂糖凝胶电泳切胶回收纯化并测定浓度后, 将 AC 库和 GT 库扩增出的序列按摩尔比 1:1 混合, 得到总 cDNA 突变测序文库。将总 DNA 突变测序文库和总 cDNA 突变测序文库进行 PE150 高通量测序, 测序数据量为 3 Gb, 测序文库的进一步构建及上机测序由武汉希望组生物科技有限公司完成。

1.8 数据处理与统计

1.8.1 数据预处理与筛选

高通量测序完成后返回的结果为原始数据, 该数据由公司经过 Q30 质量控制且已去掉接头序列。使用 Pear (v0.9.6) 软件对双端测序结果进行拼接, 然后使用 Seqkit (v2.3.0) 软件进行后续处理。使用 amplicon 指令提取需要的序列区域, 通过 rmdup 指令去除同一 read 的多重扩增子, 通过 seq-m 18-M 18 指令将符合突变序列的片段筛选出来。

1.8.2 数据统计

通过 Python 脚本分别统计总 DNA 和总 cDNA 库中出现的所有 5'-UTR 变体的种类数(即实际库容)及其 reads 数。为了比较不同 5'-UTR 变体对 mRNA 丰度的调控效果, 通过统计各 5'-UTR 变体相对每条 DNA 对应的 RNA 数量即 RNA 相对丰度, 将各 5'-UTR 变体对应的 RNA 相对丰度进行对比以估计其调控下目的基因的 mRNA 丰度差异。具体计算方式如公式(3)所示。

$$\text{RNA relative abundance} = \frac{\frac{\text{RNA reads}}{\text{Total RNA reads}}}{\frac{\text{DNA reads}}{\text{Total DNA reads}}} \quad (3)$$

式中: RNA reads 为 5'-UTR 变体在总 cDNA 库中的 reads 数; total RNA reads 为总 cDNA 库中所有 5'-UTR 变体的总 reads 数; DNA reads 为 5'-UTR 变体在总 DNA 库中的 reads 数; total DNA reads 为总 DNA 库中所有 5'-UTR 变体的总 reads 数。

使用 seqkit fx2tab 指令统计序列突变区域的碱基组成; 使用 RNAfold (v2.5.1) 软件统计

5'-UTR 序列折叠形成局部二级结构的最小自由能 ΔG (单位为 kcal/mol, 统计区域为从转录起始位点到起始密码子前 1 位碱基), 其值越小说明局部二级结构越稳定; 使用 RNAcofold (v2.5.1) 软件统计 SD 序列与 anti-SD 序列(5'-ACCUCUUA-3')结合亲和力, 计算结果为 2 条序列结合的最小自由能 ΔG (单位为 kcal/mol), 其值越小说明结合越稳定; 使用 weblogo (3.7.12) 软件制作序列的 seqlogo 图。

2 结果与分析

2.1 通过高通量测序估计不同 5'-UTR 调控下目的基因的 mRNA 丰度差异

为了系统探索 5'-UTR 与 mRNA 丰度的关系, 本研究构建了 5'-UTR 文库(图 2)。基于序列 5'-CACAAGGAG-3', 在核心 SD 序列(GGAGG)边缘和核心区域分别设置了 4 个和 1 个突变位点(分别为 5'-NNNAAGGAN-3'和 5'-CACAAGNAG-3')。此外, 本研究还设计了 2 种间隔区序列 5'-ACACAC-3'和 5'-GTGTGT-3', 以研究间隔区碱基组成对基因表达的影响。通过酶切和连接, 成功构建了一个理论库容为 518 个 5'-UTR 变体(SD 区域为 259 种序列, 间隔区域为 2 种序列)的突变文库。随后, 通过 DNA 测序以及由 RNA 逆转录成的 cDNA 测序, 深入分析了 5'-UTR 对 mRNA 丰度的影响。

经过对原始数据的清洗和筛选, 统计得到共 1.2×10^7 条 reads, 涵盖了 518 个 5'-UTR 变体, 平均每个 5'-UTR 变体检出数为 22 278 条 reads。根据测序所得每条 5'-UTR 变体对应 RNA 与 DNA 的 reads 数之比计算得到 RNA 相对丰度, 该文库中在不同 5'-UTR 调控下 RNA 相对丰度跨度为 7.9 倍(图 3A)。为检验该方法计算所得的 RNA 相对丰度的可信度, 在文库中根据 RNA 相对丰度从高到低选取 4 条 5'-UTR 变体用于调控下游 *egfp* 基因表达, 并通过

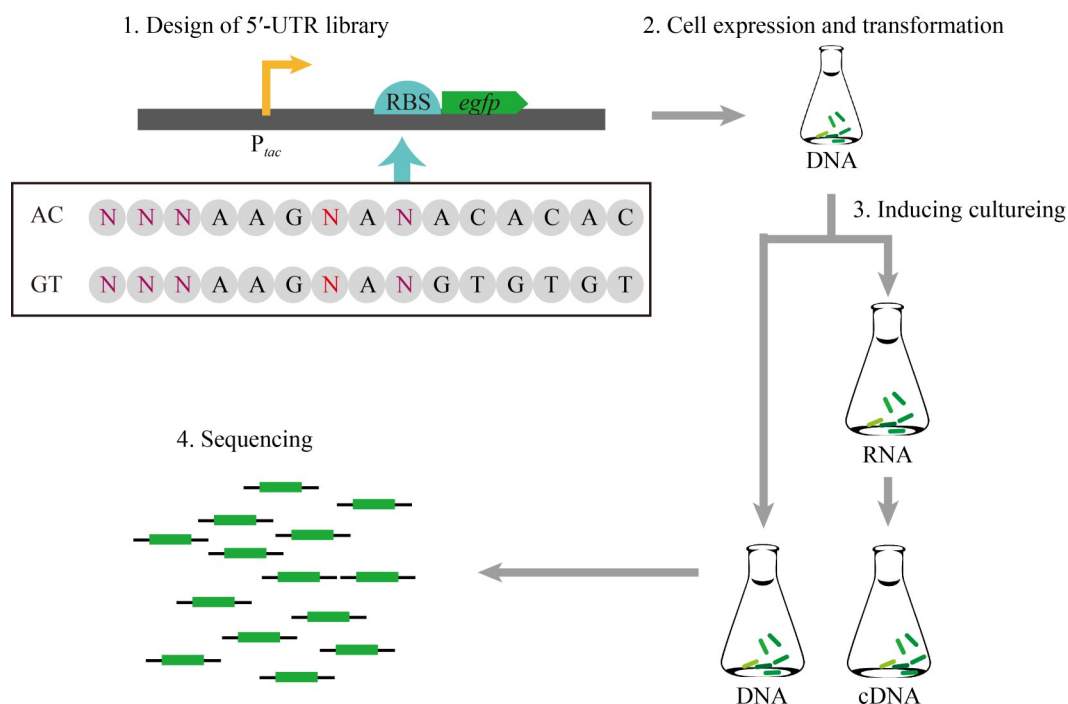


图2 5'-UTR文库的设计、构建和基于高通量测序解析各突变体调控下转录本丰度的研究流程

Figure 2 The research process of designing and constructing a 5'-UTR library and analyzing the transcript abundance regulated by each mutant based on high-throughput sequencing.

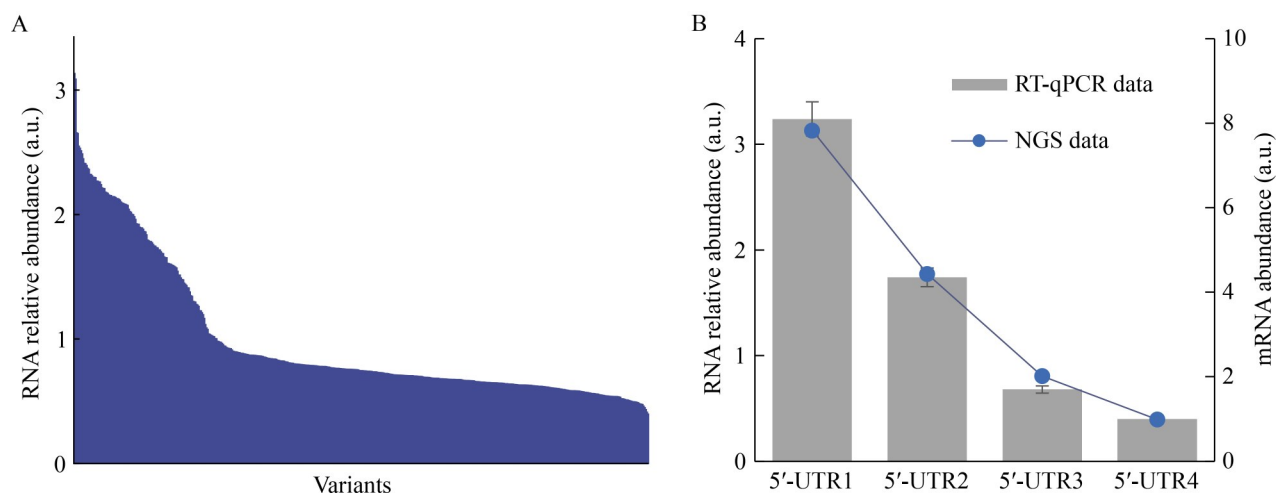


图3 利用高通量测序数据评估各突变体5'-UTR调控下的mRNA丰度。A：518条5'-UTR突变体对应的RNA相对丰度；B：文库中随机选取4条5'-UTR对应的由测序数据计算所得RNA相对丰度与RT-qPCR测得mRNA丰度对比。

Figure 3 Leveraging NGS data to estimate mRNA abundance regulated by each 5'-UTR variant. A: RNA relative abundance of 518 5'-UTR variants; B: Comparison of RNA relative abundance derived from NGS data and mRNA abundance obtained through RT-qPCR of four 5'-UTR variants randomly picked from library.

RT-qPCR 检测其 mRNA 丰度。试验结果表明, 根据测序数据计算所得的 RNA 相对丰度与试验检测得到的 mRNA 丰度在不同样本间的差异趋势与幅度一致(图 3B)。这表明通过高通量测序估计不同 5'-UTR 调控下目的基因的 mRNA 丰度差异是准确和可靠的。

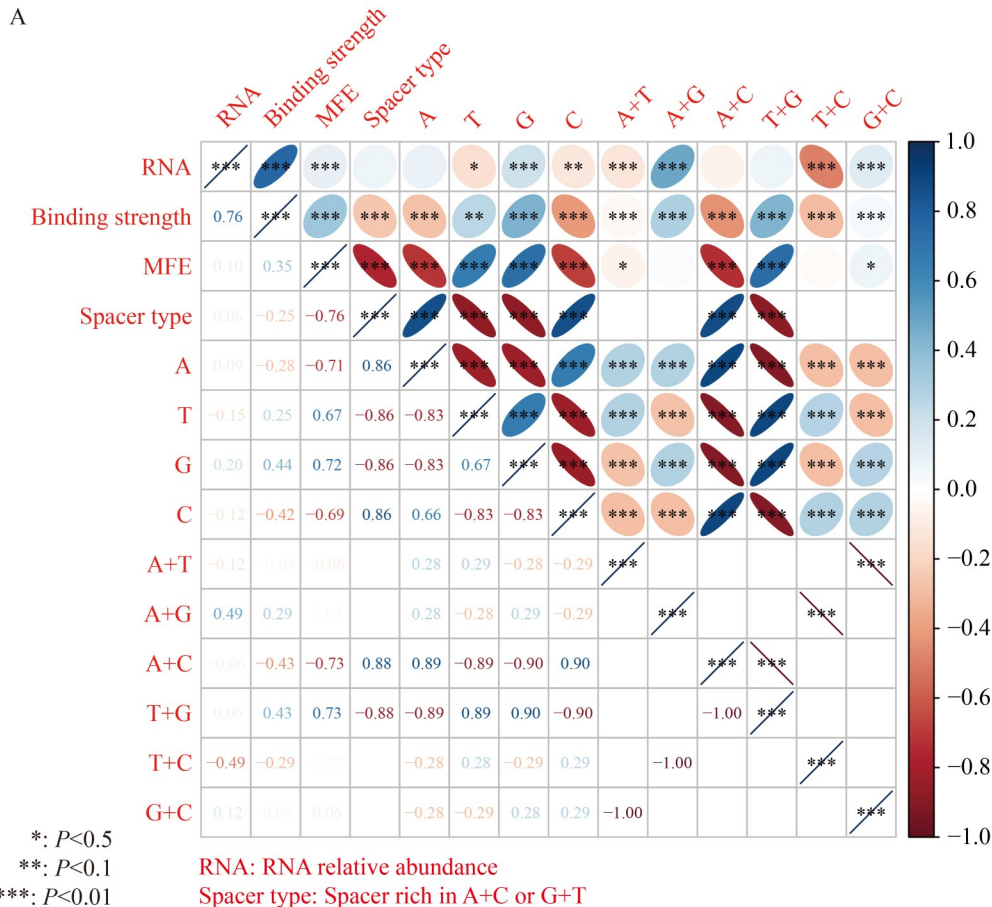
2.2 在转录后水平影响 mRNA 丰度的因素

对 5'-UTR 突变体的序列进行特征统计, 这些序列特征包括碱基组成(A、C、T、G、A+T、A+G、A+C、T+G、T+C、G+C)、5'-UTR 最小自由能[MFE, 单位为 kcal/mol, 计算公式见 1.8.2 节, 计算结果原始数据存储在国微生物科学数据中心 (http://nmdec.cn, 编号为 NMDCX0002120)]、5'-UTR 与核糖体 16S rRNA 基因(5'-ACCUCCUUA-3')的结合强度[以结合自

由能绝对值 $|\Delta G|$ 表示, 单位为 kcal/mol, 计算公式见 1.8.2 节, 计算结果原始数据存储在国微生物科学数据中心(编号为 NMDCX0002120)](以下简称结合强度)、间隔区域类型(AC 富集或 GT 富集), 并将 RNA 相对丰度与这些特征进行相关性分析(图 4A)。对 RNA 相对丰度从高到低按照最大-最小归一化后的数值(范围为 0-1)分为 5 组[高(0.8, 1.0)、较高(0.6, 0.8)、中(0.4, 0.6)、较低(0.2, 0.4)、低(0, 0.2)], 根据各组的 5'-UTR 序列制作 seqlogo 图以解析不同组间的碱基偏好性(图 4B)。

2.2.1 5'-UTR 与核糖体 16S rRNA 基因的结合强度

相关性分析结果(图 4A)显示, RNA 相对丰度与结合强度之间存在显著的强正相关($r_s=0.76$, $P<0.001$), 表明与核糖体结合强度高的 5'-UTR



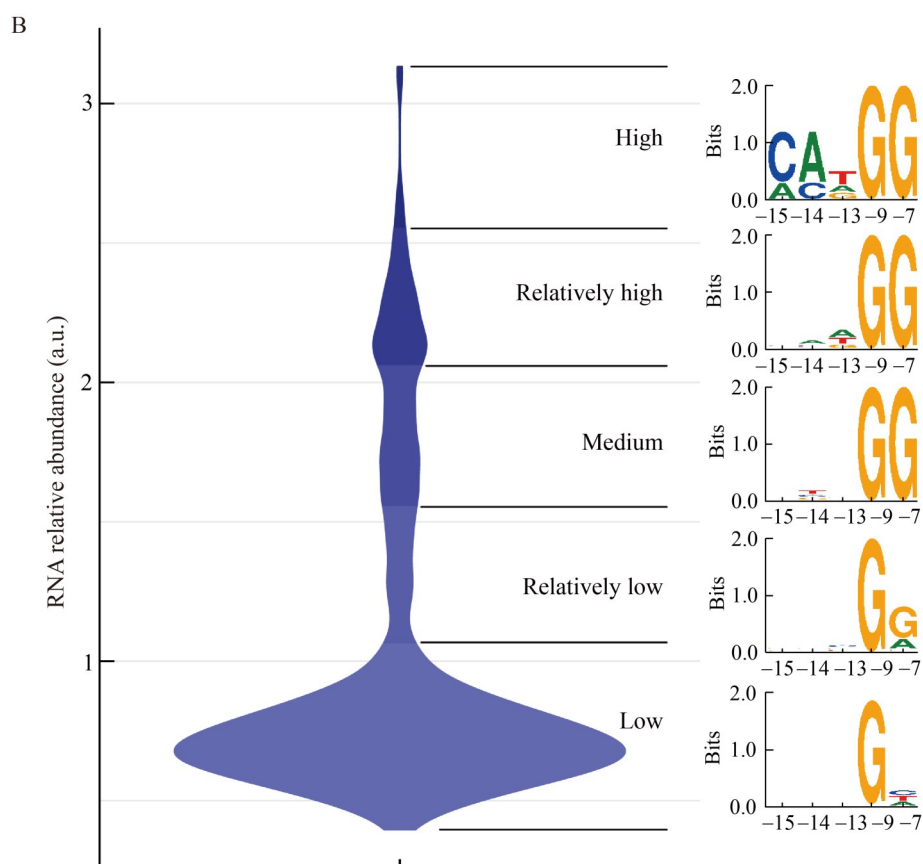


图4 5'-UTR突变体的结构序列特征对下游RNA相对丰度的影响因素分析。A: RNA相对丰度与5'-UTR突变体各序列及结构特征的相关性分析(蓝色代表正相关,红色代表负相关,颜色越深表示相关性越强); B: 按照RNA相对丰度对所有5'-UTR突变体进行分组,各组中所含5'-UTR突变体的seqlogo图(seqlogo图中横坐标数字以起始密码子第一个碱基处为+1位,该处上游第一个碱基处为-1位,以此类推;图中仅展示了在不同分组间存在碱基偏好性的碱基位点)。

Figure 4 Analysis of the impact factors of 5'-UTR variant structural sequence features on downstream RNA relative abundance. A: Analysis of correlation between RNA relative abundance and 5'-UTR variants' sequence and structure features (Blue represents positive correlation and red represents negative correlation. The darker the color, the stronger the correlation); B: All 5'-UTR variants were grouped according to the RNA relative abundance, and draw seqlogo diagrams of the 5'-UTR variants contained in each group were drawn (The horizontal digit in the seqlogo diagram is +1 at the first base of the start codon, and -1 at the first base upstream of the start codon, and so on. The figure shows only base sites with base preferences between different groups).

更有利于提高 mRNA 水平。非线性回归分析(三阶多项式拟合, $R_{\text{adj}}^2=0.78$, $P<10^{-5}$) (图 5A)表明,随着结合强度的提高, RNA 相对丰度也随之增加。这可能是因为具有高结合强度的 5'-UTR 与核糖体结合更牢固,提高了 mRNA 的翻译起始速率^[25],使核糖体稳定覆盖 mRNA^[26],从而阻

碍了 RNases 的结合位点^[11],导致 RNA 相对丰度增加。然而,过高的结合强度使 mRNA 相对丰度值略有下降,这可能是由于过强的结合强度一定程度上降低了总体翻译效率^[25],反而使 mRNA 上核糖体的覆盖率有所降低^[26],进而削弱了核糖体对 mRNA 的保护作用。通过人工设

计并构建不同结合强度的 5'-UTR 用于调控下游 *egfp* 表达, 利用 RT-qPCR 检测其 mRNA 丰度, 以评估其对基因表达转录后的影响水平。试验结果显示(图 5B), 随着结合强度的增加, mRNA 丰度呈上升趋势, 但当结合强度增加到一定程度后继续增加, mRNA 丰度表现出下降趋势, 该现象与回归分析预测的结果一致。由此证明具有与核糖体中等偏高结合强度的 5'-UTR 可以使 mRNA 丰度最大化, 过高或过低会使其出现不同程度的下降。

2.2.2 核心 SD 序列的保守性

各 RNA 丰度分组的 5'-UTR 突变体的 seqlogo 图(图 4B)显示, 高、较高和中 RNA 丰度分组的 5'-UTR 突变体在-7 位处均为碱基 G, 具有高保守性; 而较低和低 RNA 丰度分组则在该处表现出明显的碱基偏好性下降。此外, 低 RNA 丰度

分组的 5'-UTR 突变体在-9 位碱基处的碱基保守性也存在下降现象。-9 和-7 位碱基都位于核心 SD 序列(core SD, cSD)中。根据核心 SD 序列 GGAGG^[27]的保守类型将 206 个 5'-UTR 突变体分为 3 组: GGAGG (完全保守)、GGAHG 和 GHAGG (H 为简并碱基, 代表 A/T/C)。从不同分组对应的 RNA 相对丰度分布可以看出(图 6A), 当核心 SD 序列不完全保守时, RNA 相对丰度表现出显著的下降现象。通过人工设计并构建具有不同核心 SD 序列保守性的 5'-UTR 用于调控下游 *egfp* 表达, 利用 RT-qPCR 检测其 mRNA 丰度(图 6B)。试验结果显示, 当核心 SD 序列不完全保守时, mRNA 丰度明显下降, 表明核心 SD 序列的保守性会对 mRNA 丰度产生巨大影响。进一步分析其原因, 发现核心 SD 序列不完全保守的 5'-UTR 与核糖体的结合强度相比于完

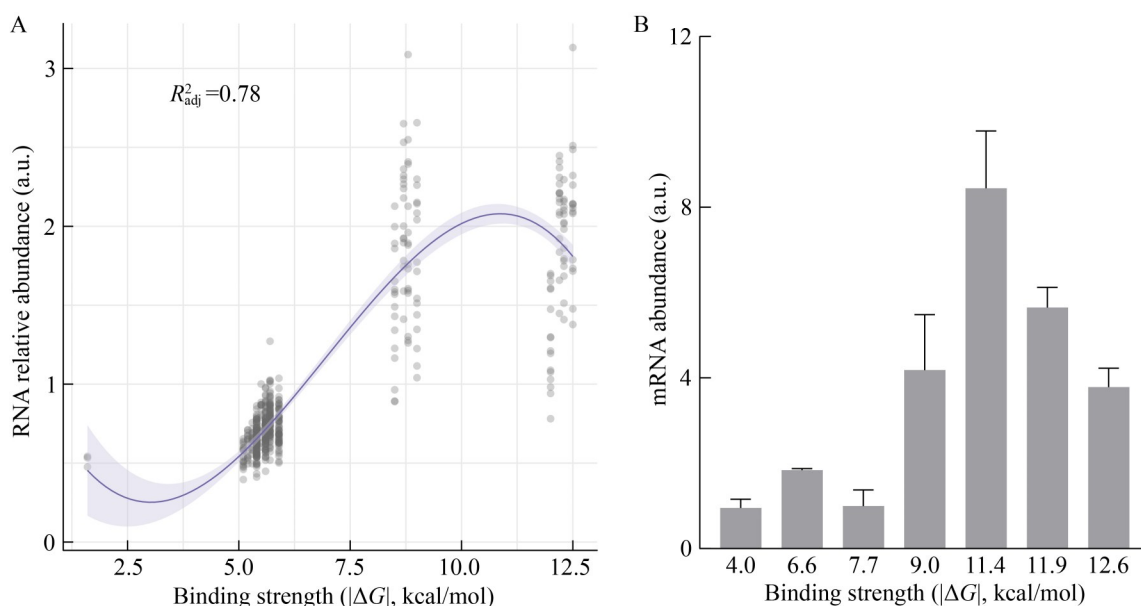


图5 5'-UTR和核糖体之间的结合强度与下游RNA相对丰度的关系。A: 5'-UTR和核糖体的结合强度与RNA相对丰度的散点图[图中曲线为三阶多项式拟合曲线($R^2_{adj}=0.78$, $P<10^{-5}$)]; B: 通过RT-qPCR检测不同结合强度5'-UTR调控下*egfp*的mRNA丰度。

Figure 5 The relationship between binding strength of 5'-UTR with ribosome and RNA relative abundance. A: Scatter plot of the binding strength of 5'-UTR with ribosome and the RNA relative abundance, the curve in the figure is a third-order polynomial fitting curve ($R^2_{adj}=0.78$, $P<10^{-5}$); B: The mRNA abundance of *egfp* regulated by 5'-UTR with different binding strength was detected by RT-qPCR.

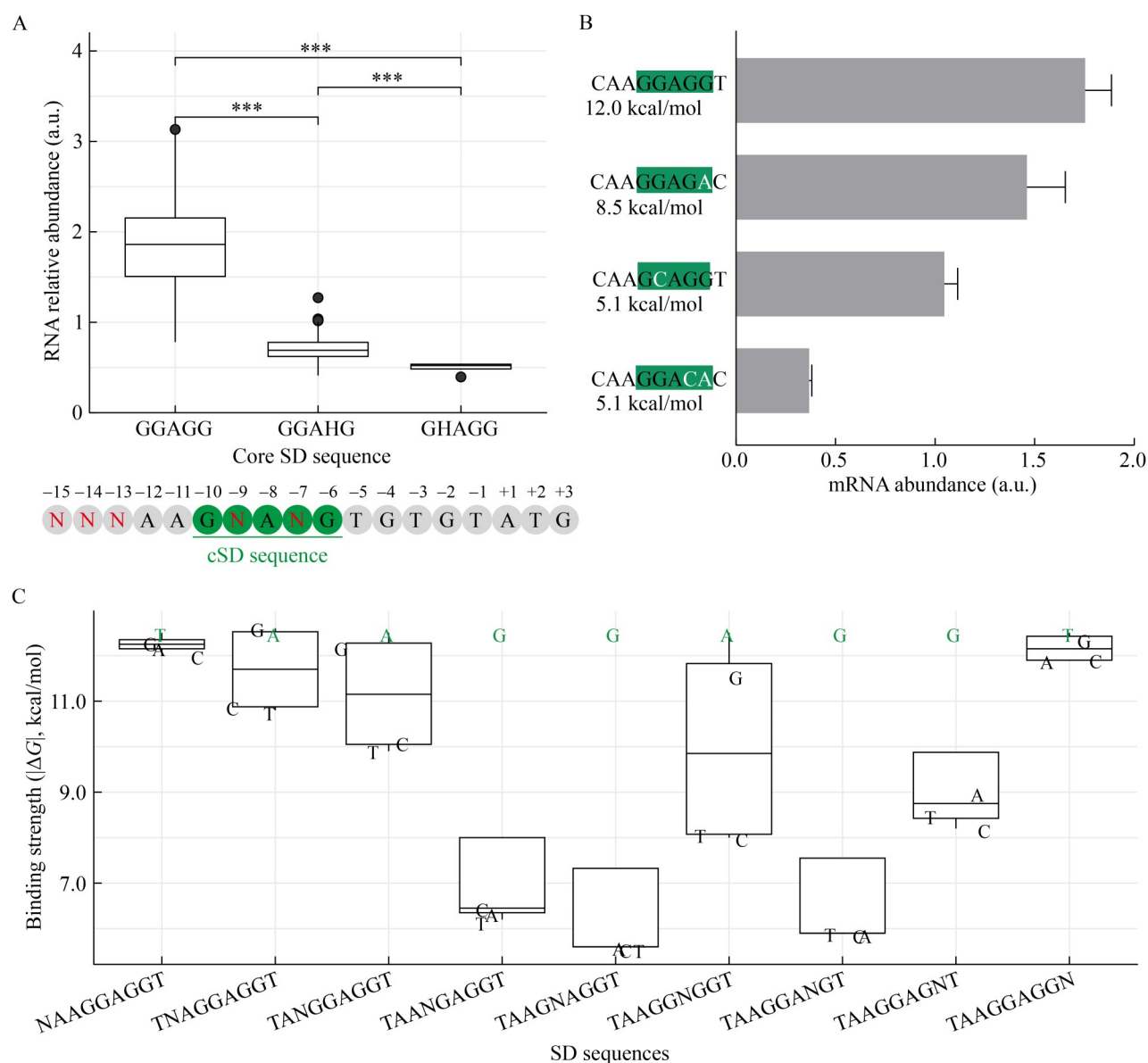


图6 5'-UTR中核心SD序列保守性与其下游RNA相对丰度的关系。 A: 具有不同核心SD序列保守性的5'-UTR突变体对应的RNA相对丰度分布; B: 通过RT-qPCR检测不同核心SD保守性的5'-UTR调控下*egfp*的mRNA丰度(纵坐标中序列下方的值表示该SD序列与核糖体的结合强度); C: 各位置不同碱基SD序列的结合强度分布, 绿色字母表示该位置对应完整SD序列的碱基。

Figure 6 The relationship between core SD sequence conservation in the 5'-UTR and the relative abundance of downstream RNA. A: RNA relative abundance distribution corresponding to 5'-UTR variants with different core SD sequence conservation; B: The mRNA abundance of *egfp* under the 5'-UTR regulation of different core SD conservation was detected by RT-qPCR (The value below the sequence in the ordinate indicates the binding strength of the SD sequence with the ribosome); C: The binding strength distribution of SD sequences of different bases at each position, and the green letter indicates the base corresponding to the complete SD sequence at this position.

全保守的 5'-UTR 显著减弱。在与 anti-SD 完整互补配对的 SD 序列(5'-TAAGGAGGT-3', 称之为完整 SD 序列)基础上, 各个碱基处分别设置 A、T、C 和 G 四种碱基, 按照引入的位置分组(例如 NAAGGAGGT 组表示在第一位设置 4 种碱基)。计算并统计各组内 SD 序列的结合强度分布(图 6C), 发现当第 4-8 位的碱基与完整 SD 序列不同时, 即核心 SD 序列不完全保守时, 结合强度出现显著下降。由此可见保证核心 SD 序列的完全保守对于目的基因的高水平表达具有重要作用。

2.2.3 5'-UTR 局部二级结构

各 RNA 丰度分组的 5'-UTR 突变体的 seqlogo 图(图 4B)显示, 高 RNA 丰度分组的 5'-UTR 突变体相比其他组在-15 和-14 位碱基处存在明显的碱基偏好性, -13 位碱基处也表现出略微的碱基偏好性。分别对-15、-14 和-13 位碱基处不同的碱基组成的 5'-UTR 突变体统计其 RNA 相对丰度分布(图 7A), 发现在-15 和-14 位碱基处具有碱基偏好性: 当 5'-UTR 突变体-15 位碱基为 A 或 C 时, RNA 相对丰度整体比碱基为 G 或 T 更高; 而-14 位碱基为 A 时相对丰度整体比碱基为 C、G 和 T 更高; -13 位碱基为 C 时, RNA 相对丰度整体比碱基为 A、G 和 T 更低。相关性分析结果表明(图 4A), RNA 相对丰度与 MFE 存在显著的弱正相关关系($r_s=0.10$, $P<0.001$), MFE 为 5'-UTR 最小自由能, 其数值越小表明在 5'-UTR 处越有可能形成稳定的二级结构^[28]。因此, 推测-15 和-14 位碱基可能是通过改变 5'-UTR 的局部二级结构从而影响 mRNA 丰度。将 518 个 5'-UTR 突变体的最小自由能与结合强度结合统计 RNA 相对丰度的分布(图 7B), 发现当结合强度一定时, 最小自由能越大, RNA 相对丰度越高。对本研究中通过 RT-qPCR 检测的所有 5'-UTR 调控下 *egfp* 表达的 mRNA 丰度数据进行同样的统计, 结果表现出同样的规律(图 7C)。由此推测, 在相似的结合强度下, 5'-UTR 最小自由能越大越不容易形成稳定的局部二级

结构。然而 mRNA 的二级结构不利于核糖体对起始位点的识别^[6,8], 因此减少 5'-UTR 二级结构的形成增加了 RBS 对核糖体的可及性^[29], 可以有效提高核糖体与 mRNA 结合并提高基因表达的效率。

3 讨论与结论

本研究通过对突变文库所在区域建立测序文库, 利用高通量测序检测目的基因表达的 DNA 和 RNA (逆转录为 cDNA) 片段数量, 通过计算 RNA 的 reads 数与 DNA 的 reads 数比值得到 RNA 相对丰度。经过试验验证, 这一方法是较为准确和有效的。通过该方法, 可以在一次试验中获取大量有关突变体的表达数据, 并以单碱基分辨率从突变体序列中提取多维度特征, 满足了对目标序列构效关系进行详尽系统分析的要求。本研究中构建的文库理论库容为 518 个突变体, 高通量测序 3 Gb 数据检出的实际库容达到了全覆盖, 平均每个突变体序列检出数约为 2 万, 其中检出数最低为 1 026, 远高于大多数分析筛选阈值 $10^{[13,19,21]}$, 理论上库容设计仍有约 100 倍的提升空间。因此, 在设计突变文库时, 至少可将理论库容扩大至 16 384 个突变体(对应 7 个饱和突变碱基位点), 在扩大分析样本容量的同时保证每个突变体的测序深度以满足分析要求。

通过对 518 个 5'-UTR 突变体的序列特征与 RNA 相对丰度的相关性分析, 发现 SD 序列与核糖体的结合强度以及 5'-UTR 序列的最小自由能分别与 RNA 相对丰度呈显著的强正相关和弱正相关。然而, 结合强度与 RNA 相对丰度的正相关并非线性关系, 而是随着结合自由能绝对值的增加, RNA 相对丰度先上升后下降, 峰值出现在中等偏高结合强度的 SD 序列处。过高的结合强度会导致 RNA 相对丰度一定程度地下降。此外, 核心 SD 序列的保守性是影响 RNA 相对丰度的重要因素, 核心 SD 序列不完全保守时会造成 RNA 相对丰度的明显下降。5'-UTR 序

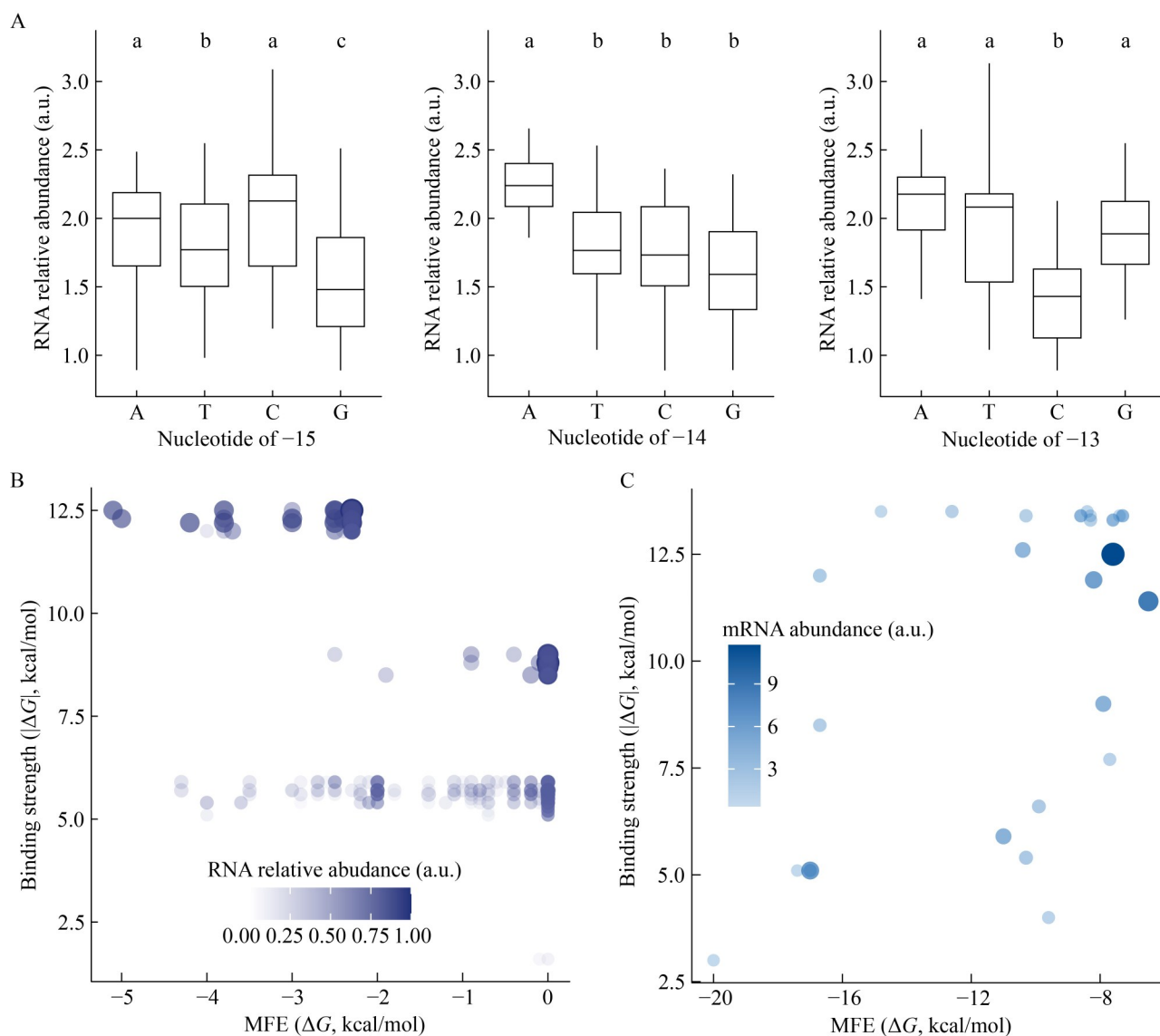


图7 5'-UTR局部二级结构与RNA相对丰度的关系。 A: 不同碱基突变位点的5'-UTR突变体对应的RNA相对丰度分布(图中不同小写字母表示各碱基分组间存在显著性差异, $P<0.05$); B: 518个5'-UTR突变体的局部最小自由能与结合强度的散点图; C: 本研究所有人工设计并构建的5'-UTR序列的局部最小自由能与结合强度的散点图。图中点越大、颜色越深, 表示对应的mRNA丰度越高。

Figure 7 The relationship between local secondary structure of 5'-UTR and RNA relative abundance. A: The RNA relative abundance distribution corresponding to different base mutation sites in 5'-UTR variants (Different lowercase letters in the figure indicate significant differences among the corresponding base groups, $P<0.05$); B: Scatter plot of the local minimum free energy and binding strength of 518 5'-UTR variants, in which larger dots and darker colors indicate higher corresponding RNA relative abundance; C: Scatter plot of the local minimum free energy and binding strength of all the 5'-UTR sequences designed and constructed artificially in this study. The larger the point and the darker the color in the diagram, the higher the corresponding mRNA abundance.

列的最小自由能表征的是 5'-UTR 局部二级结构信息, 其对 RNA 相对丰度的整体影响并不明显。然而在结合强度一定时, 5'-UTR 序列的最小自由能越大, 对应的 RNA 相对丰度越高。

以上 5'-UTR 序列各特征对 RNA 相对丰度的影响均可从核糖体对 mRNA 的保护作用角度进行解释。当核糖体在 mRNA 上覆盖的范围越大越能阻碍 RNA 降解酶对其进行降解, 从而将 mRNA 丰度维持在较高水平。SD 序列与核糖体的结合强度越强则 mRNA 招募核糖体的能力越强, 而完全保守的核心 SD 序列是保证较高结合强度的关键。然而, 过强的结合力反而导致核糖体由起始状态转变为翻译延伸状态的速率变小^[27], 导致核糖体在 mRNA 上的整体覆盖率降低。因此, 综合来看中等偏高结合强度的 SD 序列更有助于提高 mRNA 丰度, 这一现象与许多研究中观察到的现象相似, 即总体翻译效率^[26-27]和蛋白表达水平^[25]在 SD 序列与 anti-SD 之间的中等互补水平下最大化。5'-UTR 存在局部二级结构可能会影响核糖体识别 SD 序列从而不能很好地与 mRNA 结合, 因此也在一定程度上调控着 mRNA 的丰度。由于构建突变文库时未针对二级结构进行设计, 本研究并未深入探究具体二级结构特征(如茎环结构的茎长、环大小、所处位置等)对 mRNA 的转录后调控作用, 而是以 5'-UTR 序列的最小自由能整体上表征二级结构的稳定性, 并探究其与 mRNA 丰度之间的关联; 从核糖体保护机制角度看, 5'-UTR 局部二级结构的特征和所处位置, 特别是相对于 SD 序列的位置(茎邻近 SD 序列、SD 序列在茎内部、SD 序列暴露在环上)会对核糖体与 mRNA 的结合产生巨大影响^[30], 从而造成 mRNA 丰度的明显变化。

综合来看, 在构建基因表达盒或调控代谢途径时, 设计或选用与核糖体具有中等偏高结合强度的 SD 序列以及避免产生二级结构的 5'-UTR 有利于维持较高的 mRNA 丰度, 实现蛋白的高水平表达; 而缺少核心 SD 序列则会大幅

降低 mRNA 丰度, 从而下调目的基因的表达水平。本研究结果可为代谢工程及基因线路中元件的理性设计提供重要参考。

作者贡献声明

梅子轮: 实验设计和开展、数据收集和处理、文章撰写和修改; 张金鹏: 协助实验操作、参与文章讨论; 邵佳怡: 执行调研; 徐国强: 方法论; 任家卫: 参与文章讨论; 张晓梅: 监督管理; 李会: 提供资源; 史劲松: 研究指导及文章审阅; 张晓娟: 研究构思和指导、文章撰写和修改指导、获取基金; 许正宏: 提供资源及文章审阅。

作者利益冲突公开声明

作者声明不存在任何可能会影响本文所报告工作的已知经济利益或个人关系。

参考文献

- [1] ALBERTS B, JOHNSON A, LEWIS J, RAFF M, ROBERTS K, WALTER P. Molecular Biology of the Cell[M]. 4th edition. New York: Garland Science, 2002: 423-424.
- [2] CHAE TU, CHOI SY, KIM JW, KO YS, LEE SY. Recent advances in systems metabolic engineering tools and strategies[J]. Current Opinion in Biotechnology, 2017, 47: 67-82.
- [3] 戴冠苹, 孙涛, 苗良田, 李清艳, 肖冬光, 张学礼. RBS 文库调控重组大肠杆菌 β -胡萝卜素合成途径关键基因提高 β -胡萝卜素合成能力[J]. 生物工程学报, 2014, 30(8): 1193-1203.
DAI GP, SUN T, MIAO LT, LI QY, XIAO DG, ZHANG XL. Modulating expression of key genes within β -carotene synthetic pathway in recombinant *Escherichia coli* with RBS library to improve β -carotene production[J]. Chinese Journal of Biotechnology, 2014, 30(8): 1193-1203 (in Chinese).
- [4] DING NN, YUAN ZQ, ZHANG XJ, CHEN J, ZHOU SH, DENG Y. Programmable cross-ribosome-binding sites to fine-tune the dynamic range of transcription factor-based biosensor[J]. Nucleic Acids Research, 2020, 48(18): 10602-10613.
- [5] ZELCBUCH L, ANTONOVSKY N, BAR-EVEN A, LEVIN-KARP A, BARENHOLZ U, DAYAGI M, LIEBERMEISTER W, FLAMHOLZ A, NOOR E, AMRAM S, BRANDIS A, BAREIA T, YOFE I, JUBRAN H, MILO R. Spanning high-dimensional expression space using ribosome-binding site

- combinatorics[J]. *Nucleic Acids Research*, 2013, 41(9): e98.
- [6] SHINE J, DALGARNO L. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1974, 71(4): 1342-1346.
- [7] CHEN H, BJERKNES M, KUMAR R, JAY E. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs[J]. *Nucleic Acids Research*, 1994, 22(23): 4953-4957.
- [8] STEITZ JA, JAKES K. How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1975, 72(12): 4734-4738.
- [9] BARRICK D, VILLANUEBA K, CHILDS J, KALIL R, SCHNEIDER TD, LAWRENCE CE, GOLD L, STORMO GD. Quantitative analysis of ribosome binding sites in *E. coli*[J]. *Nucleic Acids Research*, 1994, 22(7): 1287-1295.
- [10] DUAN YT, ZHANG XJ, ZHAI WJ, ZHANG JP, ZHANG XM, XU GQ, LI H, DENG ZH, SHI JS, XU ZH. Deciphering the rules of ribosome binding site differentiation in context dependence[J]. *ACS Synthetic Biology*, 2022, 11(8): 2726-2740.
- [11] CETNAR DP, SALIS HM. Systematic quantification of sequence and structural determinants controlling mRNA stability in bacterial operons[J]. *ACS Synthetic Biology*, 2021, 10(2): 318-332.
- [12] BALAKRISHNAN R, MORI M, SEGOTA I, ZHANG ZG, AEBERSOLD R, LUDWIG C, HWA T. Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria[J]. *Science*, 2022, 378(6624): eabk2066.
- [13] KOSURI S, GOODMAN DB, CAMBRAY G, MUTALIK VK, GAO Y, ARKIN AP, ENDY D, CHURCH GM. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110(34): 14024-14029.
- [14] CARRIER TA, KEASLING JD. Controlling messenger RNA stability in bacteria: strategies for engineering gene expression[J]. *Biotechnology Progress*, 1997, 13(6): 699-708.
- [15] PEDERSEN M, NISSEN S, MITARAI N, SVENNINGSEN SL, SNEPPEN K, PEDERSEN S. The functional half-life of an mRNA depends on the ribosome spacing in an early coding region[J]. *Journal of Molecular Biology*, 2011, 407(1): 35-44.
- [16] RICHARDS J, LUCIANO DJ, BELASCO JG. Influence of translation on RppH-dependent mRNA degradation in *Escherichia coli*[J]. *Molecular Microbiology*, 2012, 86(5): 1063-1072.
- [17] CARRIER TA, KEASLING JD. Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*[J]. *Biotechnology Progress*, 1999, 15(1): 58-64.
- [18] DEANA A, BELASCO JG. Lost in translation: the influence of ribosomes on bacterial mRNA decay[J]. *Genes & Development*, 2005, 19(21): 2526-2533.
- [19] CAMBRAY G, GUIMARAES JC, ARKIN AP. Evaluation of 244, 000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*[J]. *Nature Biotechnology*, 2018, 36(10): 1005-1015.
- [20] ESPAH BORUJENI A, SALIS HM. Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism[J]. *Journal of the American Chemical Society*, 2016, 138(22): 7016-7023.
- [21] PATWARDHAN RP, LEE C, LITVIN O, YOUNG DL, PE'ER D, SHENDURE J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis[J]. *Nature Biotechnology*, 2009, 27(12): 1173-1175.
- [22] RINGQUIST S, SHINEDLING S, BARRICK D, GREEN L, BINKLEY J, STORMO GD, GOLD L. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site[J]. *Molecular Microbiology*, 1992, 6(9): 1219-1229.
- [23] ENGLER C, KANDZIA R, MARILLONNET S. A one pot, one step, precision cloning method with high throughput capability[J]. *PLoS One*, 2008, 3(11): e3647.
- [24] LIVAK KJ, SCHMITTGEN TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ method[J]. *Methods*, 2001, 25(4): 402-408.
- [25] VIMBERG V, TATS A, REMM M, TENSON T. Translation initiation region sequence preferences in *Escherichia coli*[J]. *BMC Molecular Biology*, 2007, 8: 100.
- [26] HOCKENBERRY AJ, PAH AR, JEWETT MC, AMARAL LAN. Leveraging genome-wide datasets to quantify the functional role of the anti-Shine-Dalgarno sequence in regulating translation efficiency[J]. *Open Biology*, 2017, 7(1): 160239.
- [27] WEI YL, SILKE JR, XIA XH. Elucidating the 16S rRNA 3' boundaries and defining optimal SD/aSD pairing in *Escherichia coli* and *Bacillus subtilis* using RNA-Seq data[J]. *Scientific Reports*, 2017, 7(1): 17639.
- [28] LORENZ R, BERNHART SH, HÖNER ZU SIEDERDISSEN C, TAHER H, FLAMM C, STADLER PF, HOFACKER IL. ViennaRNA package 2.0[J]. *Algorithms for Molecular Biology*, 2011, 6: 26.
- [29] DE SMIT MH, van DUIN J. Translational initiation on structured messengers. another role for the Shine-Dalgarno interaction[J]. *Journal of Molecular Biology*, 1994, 235(1): 173-184.
- [30] KOZAK M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes[J]. *Gene*, 2005, 361: 13-37.